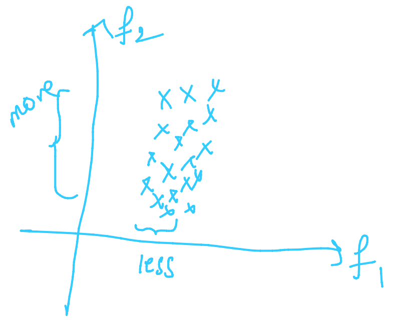
PCA Intro
25 June 2021 18:46

PRINCIPAL COMPONENT ANALYSIS.

• (Dimensionality Reduction technique).

If we have d-dimensional data & we need to reduce this to d' $(d' < d)$; then we can use PCA.



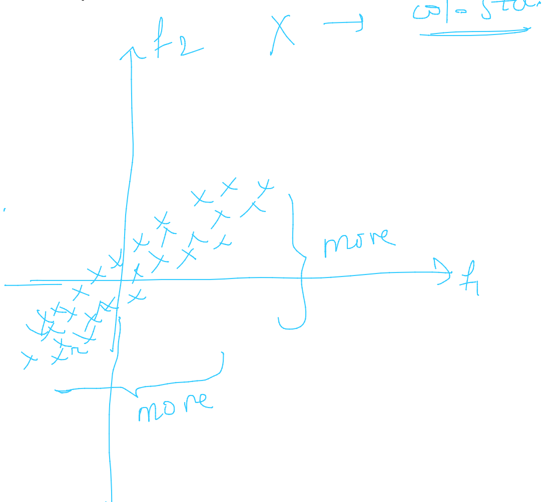
⇒ * f_1 has less variability, hence it has less info.

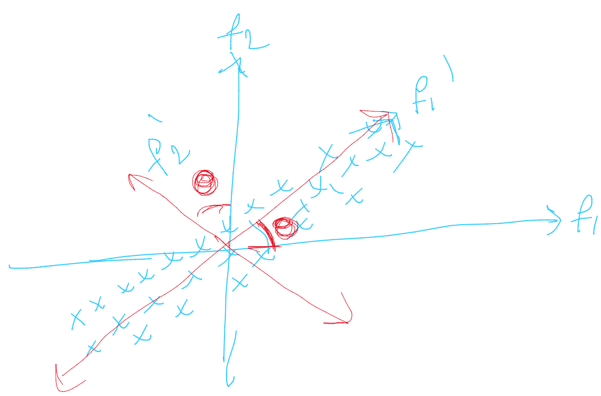
* f_2 has more variability, hence it has more information.

Hence if we have to convert this 2-D data to 1-D data then we can easily drop f_1 & project points on f_1 without loosing more information.

$$X = \begin{array}{c} f_1 \; f_2 \\ \begin{bmatrix} & \\ & \\ & \end{bmatrix}_{n\times 2} \end{array} \Longrightarrow X' = \begin{array}{c} f_2 \\ \begin{bmatrix} \\ \\ \\ \end{bmatrix}_{n\times 1} \end{array}$$

By skipping less variance feature, we could loose less.





① find f_1' & f_2' s.t. spread on $f_1' \ggg f_2'$

② Drop f_2'

③ project x_i's on f_1'

⇒ 2-D ⟶ 1-D

Main Idea

(3) project

$\Rightarrow$ 2-D $\longrightarrow$ 1-D

Main Idea

- After standardization, PCA tries to find the eigen vectors which direction of maximum variance.
- Before applying PCA, col. standardization has to be done.
  Now mean = 0 & var = 1.

Then after applying PCA, each axis is rotated to different directions.
$\Rightarrow$ co-ordinates of the points also changes. $\Rightarrow$ Variance changes

- If features are completely ~~uncorrelated~~ then PCA is useless.
- Using covariance matrix, $\Rightarrow$ Eigen vectors are P.C.s
- Eigen vectors corresponding to top k - eigen values are choosed which choose the vector which has max. variance / max eigen values.